Reliability of Marine Structures Program

PROBABILISTIC MODELS OF DYNAMIC RESPONSE AND BOOTSTRAP-BASED ESTIMATES OF EXTREMES: THE ROUTINE MAXFITS

Ron de Jong

Steven R. Winterstein

Civil Engineering Department, Stanford University

Supported by

Offshore Technology Research Center

Sandia National Laboratories

Report No. RMS–34

June 1998

Department of CIVIL ENGINEERING
STANFORD UNIVERSITY

PROBABILISTIC MODELS OF DYNAMIC RESPONSE AND BOOTSTRAP-BASED ESTIMATES OF EXTREMES: THE ROUTINE MAXFITS

Ron de Jong

Steven R. Winterstein

Civil Engineering Department, Stanford University

Supported by

Offshore Technology Research Center

Sandia National Laboratories

Report No. RMS–34

June 1998

Contents

Executive Summary

This report describes and illustrates the use of the routine MAXFITS. This routine estimates statistics of extremes corresponding to arbitrary dynamic load or response processes. It estimates statistics of extremes from limited-duration time histories, which may arise either from experimental tests or computationally expensive simulation. A wide range of statistics—e.g., mean, standard deviation, and arbitrary fractiles—can be estimated for an extreme over an arbitrary duration T. The routine also assesses, through bootstrapping methods, the statistical uncertainty associated with these extremal statistics due to the amount of data at hand. This will consistently reflect the growing uncertainty as, for example, we extrapolate to (1) increasingly high fractiles of the extreme response; or (2) increasingly long target durations T, relative to the length of the input signal.

Central to this routine is a core group of algorithms used to probabilistically model various aspects of the dynamic process of interest. The user is permitted to model either the time history itself, a set of local peaks (maxima), or a coarser set of global peaks (e.g., 5- or 10-minute maxima). A number of distribution types are included for these various purposes. For example, normal distributions and their 4-moment transformations ("Hermite") are included as likely candidates to apply directly to the process itself. Weibull models and their 3-moment distortions ("Quadratic Weibull") have been found particularly useful in modelling local peaks and ranges. Extremal, Gumbel models are also included to permit natural choices of global peaks. These algorithms build on the distribution library of the FITS routine documented in RMS Report 31 (Kashef and Winterstein, 1998).

To focus on upper tails of interest, the user can also supply an arbitrary lower-bound threshold, x_{low}, above which a shifted version of a positive random variable model—exponential, Weibull, or quadratic Weibull—is fit. In estimating the annual maximum response, the program automatically adjusts for the decreasing rate of response events as the threshold x_{low} is raised.

This program is intended to be applicable to general cases of dynamic response. A particular example shown here concerns the extreme offset statis-

tics of a floating spar buoy offshore structure. This parallels the ongoing floating structure research carried out by the Offshore Techonology Research Center, who has adopted the spar as a "theme structure" for both experimental and analytical study.

1 Introduction

1.1 Background and Motivation

This report describes and illustrates the use of the routine MAXFITS. This routine estimates statistics of extremes corresponding to arbitrary dynamic load or response processes. It estimates statistics of extremes from limited-duration time histories, which may arise either from experimental tests or computationally expensive simulation. A wide range of statistics—e.g., mean, standard deviation, and arbitrary fractiles—can be estimated for an extreme over an arbitrary duration T. The routine also assesses, through bootstrapping methods, the statistical uncertainty associated with these extremal statistics due to the amount of data at hand. This will consistently reflect the growing uncertainty as, for example, we extrapolate to (1) increasingly high fractiles of the extreme response; or (2) increasingly long target durations T (relatively to the length of the input signal.

Typical problems that motivated this study include the statistical analysis of extreme wave and wind loads/responses, based on limited data from either model or field tests. Of particular interest has been the extreme offset motions of a floating "spar buoy" offshore structure, the theme structure adopted by the Offshore Technology Research Center for both experimental and analytical study. Such motions combine lightly damped, long-period motions in both surge and pitch modes—natural periods of roughly 5min and 1.5min, respectively. In view of these long-period cycles, the amount of independent information in a 1-hour model test becomes increasingly limited. A sample problem is included here based on a (simulated) 1-hour history of spar motion, obtained from a nonlinear diffraction prediction code (Ude et al, 1995).

1.2 Problem Statement: What We Seek

In general, we focus here on the extreme value X_{max} of a random process $X(t)$, over a duration T that reflects the stationary duration of the event of interest:

$$X_{max} = \max X(t)\,; \quad 0 \leq t \leq T \tag{1}$$

Minimum values can generally be estimated in turn by replacing $X(t)$ by $-X(t)$, $1/X(t)$, or another appropriate transformation. "Two-sided" maxima, e.g. of $|X(t)|$, are less directly handled unless symmetry arguments can be applied; e.g., treating $\max|X|$ over duration T as statistically equivalent to $\max X$ over duration $2T$.

Because X_{max} will vary in a random fashion over various histories of duration T, we seek various statistics of X_{max}. A first central measure is given by its mean value, $\mu_{X_{max}}$. If we supplement this by its standard deviation, $\sigma_{X_{max}}$. we have sufficient information to fit a fairly general, two-parameter distribution function to X_{max}. Alternatively, we may directly seek various fractiles, x_p, defined so that

$$P[X_{max} \leq x_p] = p \text{ for fixed p} \tag{2}$$

Here the probability level, p, is specified and the consistent fractile x_p is sought. For example, with p=0.50, $x_{.50}$ is a representative or "median" level, which is equally likely to be exceeded or not in a given duration T. Upper fractiles of x may be useful to report to cover response variability; for example, it has recently been suggested that the p=.85 or .90- fractile response maximum provides a useful estimate, when used with the 100-year seastate, to predict the 100-year response (Engebretsen and Winterstein, 1998; Winterstein and Engebretsen, 1998).

Finally, we also may invert Eq. 1; i.e., seek the probability level p for which a specified x is not exceeded:

$$P[X_{max} \leq x] = p \text{ for fixed x} \tag{3}$$

The `MAXFITS` routine permits the user to obtain statistics in the form of either Eq. 2 or Eq. 3.

1.3 Problem Methodology: What We Model

We may seek to model a random process at a variety of different time scales. We begin here at the finest time scale, and proceed to increasingly global time scales.

Model of the entire process, $X(t)$. At the finest time scale, we may seek to model the cumulative distribution function (CDF) $F_x(x)$ of the random process $x(t)$ selected at arbitrary time t:

$$F_x(x) = \mathrm{P}[X(t) \leq x] \tag{4}$$

In the most common case $X(t)$ is assumed Gaussian, in which case $F_X(x)$ can be evaluated numerically in terms of only the mean μ_X and standard deviation σ_X of the process $X(t)$:

$$F_X(x) = \Phi\left(\frac{x - \mu_X}{\sigma_X}\right) \tag{5}$$

in which $\Phi(u)$ is the standard normal distribution function.

Model of local peaks, Y. We may instead choose to ignore all points of the time history except its local peaks, typically defined as the largest peak per upcrossing of the mean level. For a narrow-band normal process, this results in a *Rayleigh* distribution for Y, which again depends only the mean μ_X and standard deviation σ_X:

$$F_Y(y) = 1 - \exp\left[-\frac{(x - \mu_x)^2}{2\sigma_x^2}\right] \tag{6}$$

for $y \geq 0$ only.

Model of global peaks, Z. Finally, we may instead choose the maximum value Z over a still coarser time scale, comprising multiple peaks (e.g., 10-minute maxima, 1-hour maxima). As when proceeding from the process to local peaks, this step has the advantage of focusing more locally on the upper tail of interest, and the corresponding disadvantage of using less detailed information about the time history.

In generally, the distribution function of Z is commonly estimated from that of Y as follows:

$$F_Z(z) = [F_Y(z)]^N \tag{7}$$

in which N here is the number of local peaks (Y values) within the duration over which Z extends (again, 10 minutes, 1 hour, etc.) Eq. 6 assumes both that the number of peaks, N, is deterministic and that their levels are mutually independent. Neither assumption is strictly correct, but corrections generally become insignificant as we consider extremes in the upper tails of the response probability distribution.

In the Gaussian case, combining Eqs. 6–7 yields the result

$$\begin{aligned} F_Z(z) &= \left[1 - \exp\left(-\frac{(x-\mu_x)^2}{2\sigma_x^2}\right)\right]^N \\ &\approx \exp\left(-Ne^{-(x-\mu_X)^2/2\sigma_X^2}\right) \end{aligned} \tag{8}$$

The MAXFITS routine permits the user to select both which quantity is directly input—$X(t)$, Y, or Z—and also to choose which quantity is to be probabilistically modelled: either Y or Z (although, as noted below, in a particular distribution case (IDIST=11), a distribution of Y is assigned based on the statistics of X).

The various distributions available within MAXFITS are described in subsections that follow. Once estimated, $F_Y(y)$ or $F_Z(z)$ can be used to estimate the distribution of X_{max} in Eq. 1, in a manner analogous to Eq. 7:

$$\begin{aligned} F_{X_{max}}(x) = P[X_{max} \le x] &= [F_Y(x)]^{N_Y} && (9) \\ &= [F_Z(x)]^{N_Z} && (10) \end{aligned}$$

If F_Y has been fit we use Eq. 9, in which N_Y is the number of local peaks expected in time T. If F_Z has instead been fit we use Eq. 10, in which N_Z is the number of global peaks (e.g., number of 10-minute or 1-hour segments) in time T.

The mean and standard deviation, $\mu_{X_{max}}$ and $\sigma_{X_{max}}$, corresponding to the distribution of X_{max} given above is found in MAXFITS by numerically integration, using Gaussian quadrature procedures.

1.4 Uncertainty Estimates through Bootstrapping

Finally, bootstrapping methods (e.g., Efron and Tibshirani, 1993) are used here to estimate the statistical uncertainty associated with any/all of our estimated statistics of X_{max}. The method is conceptually straightforward, generating multiple "equally likely" data sets by simulating, with replacement, from the original data set. Thus some of the data values will be repeated multiple times, while others will be omitted, in any single bootstrap sample (which is of the same size as the original data set). The same estimation procedure performed for the original data set is repeated for each of the bootstrapped samples, and the net statistics on the results are collected and reported.

The bootstrap method is "non-parametric" by definition, in that it operates with no additional information beside the actual data values. Alternative approaches might fit a parametric model, either statistical or physical, to generate additional "equally likely" samples from which to infer sampling variability levels. Such approaches may confer advantages in some cases but are generally problem-specific; the bootstrap method is adopted here primarily due to its virtue of generality.

1.5 Distribution Fitting; Relation to Other Algorithms

Central to this routine is a core group of algorithms used to probabilistically model various aspects of the dynamic process of interest noted above: the process X, its local peaks Y, or its global peaks Z. The set of distribution types available are, with the sole exception of the 4-moment Hermite model, the same as those available in the routine `FITS`, as documented in RMS Report 31 (Kashef and Winterstein, 1998). Again apart from the Hermite case, this distribution set was chosen to provide relatively robust fits, preserving two or at most three moments.

In this sense, both `FITS` and `MAXFITS` are intended to complement the previously distributed routine, `FITTING`, documented in RMS Report 14 (Win-

terstein et al, 1994). The FITTING routine implements relatively complex, four-moment distribution models, whose parameters are fit with numerical optimization routines. While these four-moment fits can be quite useful and faithful to the observed data, their complexity can make them difficult to automate within standard fitting algorithms, and repeated application over sets of bootstrapped samples. As noted above, however, we do include the 4-moment Hermite distribution as implemented in FITTING, in view of its growing use in a variety of applications.

To focus on upper tails of interest, the user can also supply an arbitrary lower-bound threshold, x_{low}, above which a shifted version of a positive random variable model—exponential, Weibull, or quadratic Weibull—is fit. (In estimating the annual maximum response, the program automatically adjusts for the decreasing rate of response events as the threshold x_{low} is raised.)

1.6 Available Distribution Types

Specific distributions currently included in MAXFITS to estimate $F_i(x)$ include the following, as catalogued by the distribution index IDIST:

IDIST=1: Normal Distribution

IDIST=2: Lognormal Distribution

IDIST=3: Exponential Distribution

IDIST=4: Weibull Distribution

IDIST=5: Gumbel Distribution

IDIST=6: Shifted Exponential Distribution

IDIST=7: Shifted Weibull Distribution

IDIST=8: Quadratic Weibull Distribution

IDIST=9: Shifted Quadratic Weibull Distribution

IDIST=10: Four-Moment Hermite Distribution

IDIST=11: Hermite Distribution Model of Peaks, based on four moments of the underlying process

The distributions IDIST=1 through 5 and 8 are all fit to statistical moments of all available data. The single-parameter exponential preserves only the mean m_x of the data, while the normal, lognormal, Weibull, and Gumbel preserve both the mean and standard deviation σ_x estimated from the data. The quadratic Weibull preserves the first three moments of the data (mean, standard deviation, and skewness). The Hermite model (IDIST=10) is perhaps the most general, seeking to preserve the first four moments of the data (mean, standard deviation, skewness, and kurtosis). The Hermite model of peaks (IDIST=11) is special, in that it takes as input the first four moments

of the underlying random process $X(t)$, and provides a consistent distribution of the local peaks Y.

Most of the one-sided distributions above (exponential, Weibull, and quadratic Weibull) are also generalized here by shifting (IDIST=6, 7, and 9). These impose a user-defined lower threshold x_{low}, ignore data below x_{low}, and fit standard exponential/Weibull/quadratic Weibull models to $x - x_{low}$ based on observed moments. These are perhaps the most relevant distributions when modelling local peaks, Y, which generally have a broadly skewed distribution away from a well-defined lower bound. (In estimating the annual maximum response, the program automatically adjusts for the decreasing rate of response events as the threshold x_{low} is raised.)

The result aims to provide the user with a suite of smooth probability models, to be fit throughout the body of the available data. It does not directly address various special topics of data fitting; e.g., selective tail fitting, fitting bimodal models to hybrid data, etc. Some of these issues can be addressed, in a limited way, through the use here of the *shifted* models (IDIST=6, 7, and 9). In this way the user can focus the distribution modelling resources on the extreme response levels of interest.

More specific tail-fitting procedures have not been given here, because optimal use of these tends may be rather problem-specific. In the same vein our extremal models are limited here to so-called "Type I" behavior, leading to (shifted) exponential distributions of peaks over a given threshold and to Gumbel distributions of annual maxima. Type II and III distributions are ill-suited to our moment-fits, due to potential moment divergence (Type II) or to the difficulty in predicting truncated distributions (Type III) from moment information.

1.7 Limitations

An important limitation is that for IDIST=11 the process $X(t)$ is assumed input, and its moments used to obtain a consistent distribution to assign to the local peaks, Y. In this case we do not permit the bootstrapping option,

as one would distort the time-scale of variation of $X(t)$ if its values were merely sampled with replacement over the time-axis.

`NMAX`, the maximum number of data, has been set to 45000. This has been set in a `PARAMETER` statement in the main driver program to `MAXFITS`. This is a rather arbitrarily selected limit, and can be reset by the user without fundamental consequence.

2 Distribution Fitting: Routines

The routine `MAXFITS` has been separated into three files containing Fortran source code: `maxf.f` contains the main program, `aux_fits.f` contains auxiliary subroutines used by `FITS`, and `aux_maxf.f` contains all additional subroutines used by `MAXFITS`.

Specifically, the fitting algorithm includes the following set of subroutines, contained in `aux_fits.f`:

`CALMOM`: Estimates the mean m_x, standard deviation σ_x, skewness α_3 and kurtosis α_4 from an input set of data. These are based on unbiased estimates of the cumulants k_1=m_x, k_2=σ_x^2, k_3=$\alpha_3\sigma_x^3$, and k_4=$(\alpha_4-3)\sigma_x^4$. If the user includes an optional lower limit x_{low}, moments of the shifted variable $(x - x_{low})^+$=$\max(0, x - x_{low})$ are estimated.

`DISPAR`: Based on the sample moments estimated in `CALMOM`, `DISPAR` seeks a consistent set of distribution parameters. The interpretation of these parameters depends on the distribution type selected by the user. Appendix A includes a complete listing of the distribution functions and their parameters.

`GETCDF`: For the user-defined distribution type with the distribution parameters from `DISPAR`, this routines estimates the cumulative distribution function value, $F(x)$=P[Outcome $< x$] for given input x value.

`FRACTL`: For the user-defined distribution type with the distribution parameters from `DISPAR`, this routines estimates the fractile x corresponding to a specified input value of the probability p=$F(x)$=P[Outcome $< x$].

`QDMOM`: Uses Gaussian quadrature to estimate the first four moment of the theoretical fitted distribution. These can be compared with the sample moments from the data, as given by `CALMOM`, to verify the accuracy of the fitted model—and in the case of the higher moments not used in the original fitting, to test its accuracy.

The routines GETCDF and FRACTL, which supply general distribution functions and their inverses, may also be useful in other stand-alone applications; e.g., to create a distribution library for standard FORM/SORM or simulation analyses (Madsen et al, 1986), or for use with new Inverse FORM algorithms (Ude and Winterstein, 1996).

The additional subroutines contained in `aux_maxf.f` are as follows:

DATAPREP: Prepares the data for the analysis. The user specifies whether the input data represent the entire process $X(t)$, the local peaks Y, or the global peaks Z. DATAPREP selects, from the input information, the appropriate data values to be retained for purposes of probabilistic modelling/fitting.

DISTINT: Finds the mean and standard deviation, $\mu_{X_{max}}$ and $\sigma_{X_{max}}$, of the maximum value X_{max} by numerical integration, using Gaussian quadrature methods.

RESAMP: Generates a new, "equally likely" dataset of the same size from the original data by sampling with replacement. This is ued to produce bootstrap estimates of the standard deviation of our estimates.

CALCRES: Handles administrative work involved with bootstrapping, such as keeping track of running sums, etc.

3 Input Format and Spar buoy Example

3.1 Data Input

The file containing data are read in free format, one datum per line. Non-numerical input is interpreted as commentary, and is ignored. The input file only contains the data that needs to be fitted. The first line does not contain the duration of the database, contrary to "FITS".

We will illustrate the use of "MAXFITS" though a simple example, which is the surge response of a spar buoy. The input is discussed in the following paragraph. The output is discussed in the next chapter. The data set analyzed here contains one hour of simulated data, from which the surge component is filtered. The natural period of the spar buoy is approximately 5 minutes in surge hence there are only 12 peaks, which should illustrate the implications of dealing with limited data.

The input file is stored in surge1.ts. The time series is plotted in fig 1.

3.2 Runtime Input: Batch Mode

We desire the following situation:

1. Results should be written to a file named weibull1.out
2. Distribution results are to be written for x (surge response) values ranging from XMIN=5 to XMAX=17.5m, at increments of DX=1m
3. The surge data is stored in the file surge1.ts
4. The user desires to fit a shifted Weibull distribution (IDIST=7) to these data. IDIST=4 should only be used if it is certain the mean of the underlying process equals 0. If this is not the case the fit should be shifted over the mean, or any other threshold if preferred.[1]
5. The user desires to determine the accuracy of the results by producing 100 bootstrap estimates of all the predictions

The type of input provided is specified by the INSWITCH variable. The available options are:

[1] Although it is inconvenient for the user to have to determine the mean of the process, there is no other method. The only way "MAXFITS" can determine the mean is if the entire process is input. In this case if the user specifies a value less than −1000 for XLOW "MAXFITS" will automatically use the mean of the process as threshold.

1. The entire process
2. The local peaks of the process
3. The global peaks of the process (the number of global peaks for equal time segments is specified with the NSEG variable)

The type of data we wish to use for the analysis is specified with the DATASWITCH variable, which has the same options as the INSWITCH variable.

The desired output can be selected with the OUTSWITCH variable, for which the user can select the following values:

1. The user inputs a lower limit for the input variable, an upper limit and a step size (XMIN, XMAX, DX). "MAXFITS" will output the probability of exceedence for each specified response. Bootstrapping will give a mean and standard deviation for the response.
2. The user inputs specific response values, by first specifying the number of inputs (NOUTPTS), and then the response fir which the probability of exceedence will be calculated. Bootstrapping will give a mean and standard deviation for the response.
3. MAXFITS determines the entire distribution of the probability of exceedence for a specified number of points. Probabilities will range from 1/N to 1-1/N. Bootstrapping will give a mean and standard deviation for the probability of exceedence.
4. The user inputs specific probability levels, by first specifying the numbers of inputs (NOUTPTS), and then the probability of exceedence for which the associated response will be calculated. Bootstrapping will give a mean and standard deviation for the probability of exceedence.

The previous options cause that the input lines will be different depending on the output specified on the second line of the batch file. Examples of input for all 4 possible output options are given. The batch file for the example is named weibull1.in, and contains the following input lines:

```
Weibull1.out: Name of output file
1 2 1        : INSWITCH,.DATASWITCH, OUTSWITCH
7.5 17.5 1. : XMIN, XMAX, DX[2]
1. 6. (NSEG): Duration of input file and target period[3], (#global peaks)
100          : Number of bootstrap samples
surge1.ts    : Name of input file
7            : Distribution type (IDIST), see Appendix A for definitions
0.41         : XLOW, shift only for shifted distributions
```

[2] Note that selecting XMIN too low, or XMAX too high may cause underflow errors

[3] Units are free, as long as they are consistent

Alternatively the following batch files can be used for OUTOPT = 2,3,4 respectively:

```
weibull1.out: Name of output file
1 2 2        : INSWITCH,.DATASWITCH, OUTSWITCH
3 (or Noutpts): NOUTPTS, no. of exceedence probabilities to be calculated
             : First fractile for which P will be calculated
15.          : Second extreme for which P will be calculated
20.          : Third extreme for which P will be calculated
  :
Xnoutpts     : Nth extreme for which "MAXFITS" will calculated the
               probability of exceedence
1. 6. (NSEG): Duration of input file and target period
100          : Number of bootstrap samples
surge1.ts    : Name of input file
7            : Distribution type (IDIST), see Appendix A for definitions
0.41         : XLOW, shift only for shifted distributions

weibull1.out: Name of output file
1 2 3        : INSWITCH,.DATASWITCH, OUTSWITCH
99           : Noutpts, the number fractiles that will be calculated equal
               probability intervals (0.01-0.99)
1. 6. (NSEG): Duration of input file and target period
100          : Number of bootstrap samples
surge1.ts    : Name of input file
7            : Distribution type (IDIST), see Appendix A for definitions
0.41         : XLOW, shift only for shifted distributions

weibull1.out: Name of output file
1 2 4        : INSWITCH,.DATASWITCH, OUTSWITCH
3 (or Noutpts): NOUTPTS, no. of probabilities for which fractiles will be
               calculated
0.01         : First probability of exceedence
0.001        : Second probability of exceedence
0.0001       : Third probability of exceedence
:
Pnoutpts     : Nth probability of exceedence for which "MAXFITS" will
               calculate the fractile
1. 6. (NSEG): Duration of input file and target period
100          : Number of bootstrap samples
surge1.ts    : Name of input file
7            : Distribution type (IDIST), see Appendix A for definitions
0.41         : XLOW, shift only for shifted distributions
```

By typing the following command:

```
maxfits < weibull1.in
```

A file named weibull1.out will be written whose content is discussed in the next section. During the execution the user will be prompted for terminal inputs. These can simply be ignored (or directed toward the null device) in this batch mode operation.

3.2 Runtime Input: Interactive Mode

If the user simply types "maxfits", he or she will prompted for each input, which is the same as what is described in the previous paragraph. The prompts are accompanied by interactive explanations that will list the options the user has. The interactive mode may be particularly useful for first-time users. (The text with the input prompts is written to the logical unit IOERR, which is set to 0 in the driver program. The user can reset this if necessary.)

The following is a screen dump of the terminal input and the user's response. Lines beginning with ">" are input prompts generated by the program. Other lines are the user's response, which should match the input given in the first batch file in the previous paragraph.

```
Weibull1.out
>
> ** ENTER THE TYPE OF DATA IN THE DATA FILE,
>    THE TYPE OF DATA TO BE USED FOR THE ANALYSIS,
>    AND THE OUTPUT SWITCH:
>
>    INSWITCH/DATASWITCH =  1 ...  POINTS OF THE PROCESS
>    INSWITCH/DATASWITCH =  2 ...  LOCAL PEAKS
>    INSWITCH/DATASWITCH =  3 ...  GLOBAL PEAKS
>    DATASWITCH >= INSWITCH
>
>    OUTSWITCH = 1 ... ENTER XMIN,XMAX,DX -> P1..PN
>    OUTSWITCH = 2 ... ENTER X1,X2,...,XN -> P1..PN
>    OUTSWITCH = 3 ... ENTER NP           -> X1..PN
>    OUTSWITCH = 4 ... ENTER P1,P2,...,PN -> X1..PN
>
>       ENTER INSWITCH,DATASWITCH,OUTSWITCH:
1 2 1
>
> ** ENTER XMIN  = MIN X VALUE AT WHICH TO OUTPUT CDF
>          XMAX  = MAX X VALUE AT WHICH TO OUTPUT CDF
>            DX  = INTERVAL OF X VALS WHERE CDF IS OUTPUT
>    ALL THREE VALUES ON SAME LINE; E.G.
>
>        0.5 10.0 0.5
>
>    GIVES OUTPUT AT 20 X VALUES FROM 0.5 TO 10.0
>
```

```
>       ENTER XMIN,XMAX,DX:
7.5 17.5 1.
>
> ** ENTER THE DURATION OF THE DATA FILE
>    AND THE TARGET DURATION FOR THE PREDICTION:
>    ASSURE TTARGET IS SUFFICIENTLY LONG
>    TO CONTAIN AT LEAST ONE CYCLE
>
>    Ttot,Ttarget:
1. 6.
>
> ** ENTER THE NUMBER OF BOOTSTRAP SAMPLES TO BE TAKEN:
>    FOR NO BOOTSTRAPPING ENTER bsN=0
>    bsN:
100
>
> ** ENTER FILENAME WHERE DATA ARE STORED,
>
>       ENTER INPUT FILENAME:
surge1.ts
>
> ** ENTER IDIST =INDEX OF DISTRIBUTION TYPE TO BE FIT
>    CURRENT OPTIONS:
>                      IDIST =  1 ...    NORMAL
>                      IDIST =  2 ...    LOGNORMAL
>                      IDIST =  3 ...    EXPONENTIAL
>                      IDIST =  4 ...    WEIBULL
>                      IDIST =  5 ...    GUMBEL
>                      IDIST =  6 ...    SHIFTED EXPONENTIAL
>                      IDIST =  7 ...    SHIFTED WEIBULL
>                      IDIST =  8 ...    QUADRATIC WEIBULL
>                      IDIST =  9 ...    SHIFTED QUADRATIC WEIBULL
>                      IDIST = 10 ...    HERMITE (PROCESS)
>                      IDIST = 11 ...    HERMITE (PEAKS)
>
>       ENTER IDIST:
7
>
>    YOU HAVE SELECTED A SHIFTED DISTRIBUTION MODEL
>
> ** ENTER XLOW  =LOWER BOUND THRESHOLD, BELOW WHICH
>                      ALL DATA WILL BE IGNORED
>
>       ENTER XLOW :
0.41
```

4 Output Format and Spar Buoy Example

Below is the output file "weibull1.out" that resulted from the manual input listed in the previous paragraph. The format is the same for all output options. Note that the output is formatted such that it can be directly plotted using "gnuplot". The lines starting with # will be treated as comments by "gnuplot".

The first section echo's the input, and how much data was actually used for the analysis.

The second section provides summary statistics for the data file considered. These include on the first line the sample moments from the data, and on the second line the standard deviation of the bootstrap predictions.

The third section gives the moments that are implied by the fitted distribution in the same way as they are given for the original data.

The fourth section reports the distribution parameters. The standard deviation of the bootstrap predictions is given on the second line. The definition of the distribution parameters is given in appendix A.

The fifth section gives the mean and standard deviation of the distribution of the extreme value in the target period. The bootstrap standard deviations of these values are reported on the second line.

The last section reports the actual distribution of the extreme value in the target period. The first column reports the fractiles that were input by the user in this case, but which could also have been calculated if the user had specified probability levels in the input (OUTOPT = 3,4). The second column reports the standard deviation of the in this case 100 predictions of the fractile. As the fractile is input here this column consists of zeros. The third column reports the probability of exceedence of the fractile. The fourth reports the bootstrap standard deviation of this probability, and indicates the accuracy of the prediction.

```
#
#               RESULTS FOR:   surge1.ts
# TIME DURATION OF DATABASE:         1.00
#                CONTAINING:         7200  POINTS OF THE PROCESS
#      TARGET TIME DURATION:         6.00
#        DIST TYPE SELECTED:   SHIFTED WEIBULL
#                 FITTED TO:           11  LOCAL PEAKS
#  NO. OF BOOTSTRAP SAMPLES:          100
```

```
#
#   ** NOTE: MOMENTS, DIST PARMS APPLY HERE TO X-XLOW; XLOW=  0.4100E+00
#
#
#     MOMENTS FROM SAMPLE DATA     ( MEAN, SIGMA, SKEWNESS, KURTOSIS)
#    data:    0.5023E+01  0.2661E+01 -0.4110E+00  0.2644E+01
#    stdv:    0.7448E+00  0.4486E+00  0.5298E+00  0.1152E+01
#
#
#
#     MOMENTS FROM FITTED DIST     ( MEAN, SIGMA, SKEWNESS, KURTOSIS)
#    data:    0.5023E+01  0.2661E+01  0.6508E+00  0.3283E+01
#    stdv:    0.7448E+00  0.4486E+00  0.3768E+00  0.6239E+00
#
#
#
#     DISTRIBUTION PARAMETERS      (SEE DOCUMENTATION FOR DEFINITION)
#    data:    0.5023E+01  0.2661E+01  0.1971E+01  0.5666E+01  0.0000E+00
#    stdv:    0.7448E+00  0.4486E+00  0.9153E+00  0.8210E+00  0.0000E+00
#
#
#               MEAN       STDV       (of MAX response in Ttarget):
#    data:     12.87       1.57
#    stdv:      1.58       0.55
#
#
#      X          stdv(X)     1-Fxmax     stdv(1-Fm)
  0.7500E+01  0.0000E+00  0.1000E+01  0.4344E-01
  0.8500E+01  0.0000E+00  0.9999E+00  0.1212E+00
  0.9500E+01  0.0000E+00  0.9956E+00  0.1930E+00
  0.1050E+02  0.0000E+00  0.9495E+00  0.2880E+00
  0.1150E+02  0.0000E+00  0.7899E+00  0.3124E+00
  0.1250E+02  0.0000E+00  0.5382E+00  0.2733E+00
  0.1350E+02  0.0000E+00  0.3037E+00  0.2119E+00
  0.1450E+02  0.0000E+00  0.1481E+00  0.1535E+00
  0.1550E+02  0.0000E+00  0.6481E-01  0.1060E+00
  0.1650E+02  0.0000E+00  0.2610E-01  0.7047E-01
  0.1750E+02  0.0000E+00  0.9807E-02  0.4539E-01
```

As the Weibull distribution only uses two parameters only the first two statistical moments can be reproduced by the fitted distribution. The skewness and kurtosis differ somewhat.

The original data, and the Weibull-fit which, is similar to the output of "FITS", are shown in figure 2. The two extra lines reflect the result if the fit had been biased 10% upward or downward, giving a feel for how well the data points are matched by the model.

Figure 6 shows the distribution of the 6-hour extreme surge response produced by "MAXFITS", and lines reflecting this value plus and minus two bootstrap standard deviations, which would be the 95% confidence interval if we assume the distribution of our predicted fractiles to be normally distributed. This distribution has the previous distribution as underlying distribution, which it raises to the power of the number of

peaks in 6 hours (66 in this case). The graph clearly shows that the accuracy becomes increasingly less for lower probabilities of exceedence. If one would like to estimate the 85% fractile, which is now proposed for some long-term analyses, it would have a standard deviation of 4.27. The alternative would be the mean 6-hour extreme response, which has a standard deviation of 1.58. The CoV's are respectively 0.29 and 0.12. In order to achieve the same level of accuracy it would require use to have $(0.29/0.12)^2$ = 5.8 times as much data.

It is important to note the effect of the amount of data that is used for the fit on the accuracy of our predictions. Figures 6,7,8,and 9 illustrate this. They show the results for the different components and the total of the response (horizontal offset at 54m MWL) of a spar buoy. The number of peaks, the mean of the distribution of the 6-hourly extreme, and its standard deviation are reported in the following table.

	Estimate	Stdv	CoV	Peaks
surge	12.87	1.58	0.12	11
pitch	10.98	0.87	0.08	53
wave freq.	35.11	2.9	0.08	246
total	11.28	0.5	0.04	136

The standard deviations are consistent with the graphs, which clearly show that less peaks give a larger bootstrap standard deviation, and therefore less accurate results. The underlying distributions and the original data are plotted in figures 2,3,4, and 5 to demonstrate the quality of the fit.

5 Problems / Pitfalls

5.1 Underflow errors

Enter a too low XMIN:
Xmin is way below the data and therefore a very low probability associated with it. This can cause underflow errors.

Enter a too high XMAX:
MAXFITS has to extrapolate very far to very low probabilities of exceedence, which may cause underflows

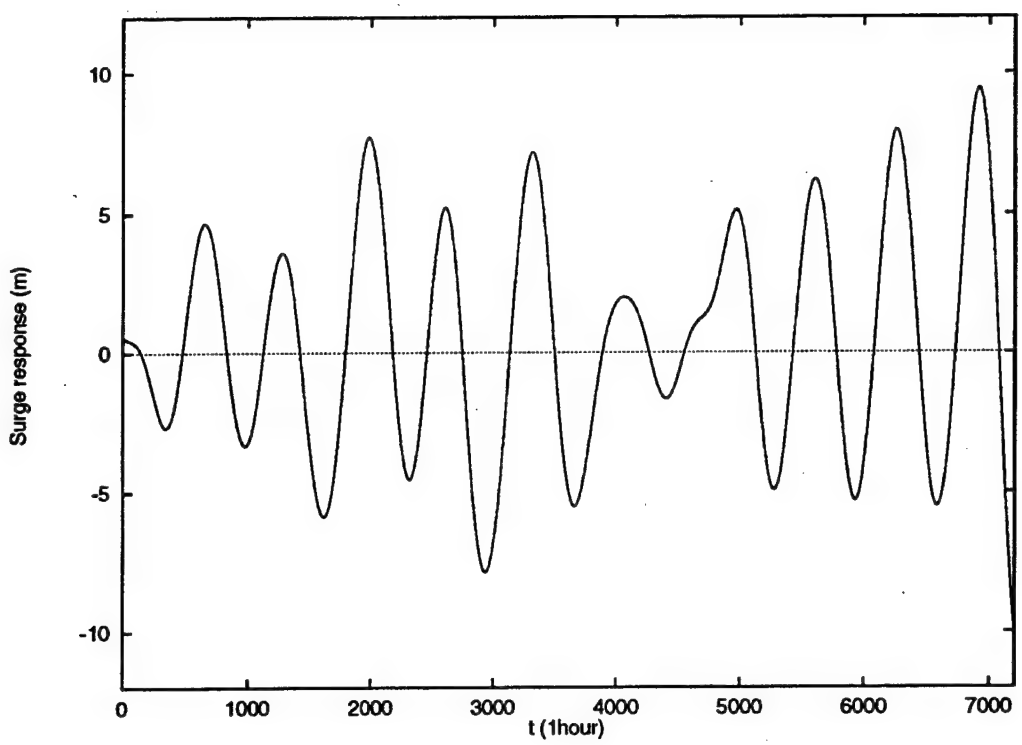

Figure 1: Simulated time series of surge motion.

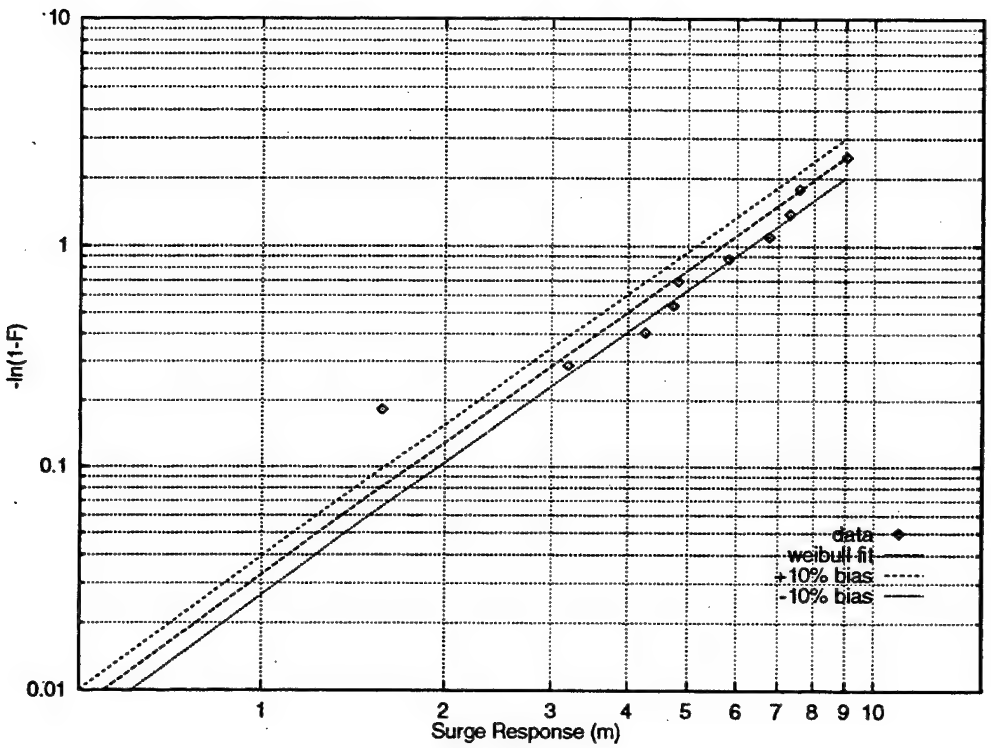

Figure 2: Distribution of local peaks, surge component.

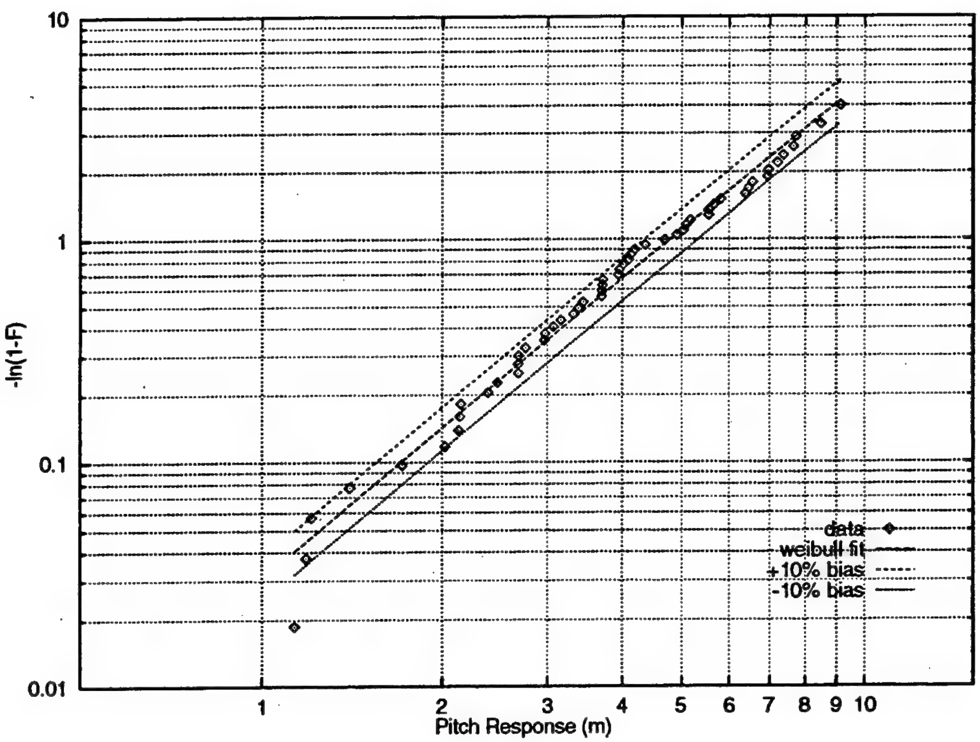

Figure 3: Distribution of local peaks, pitch component.

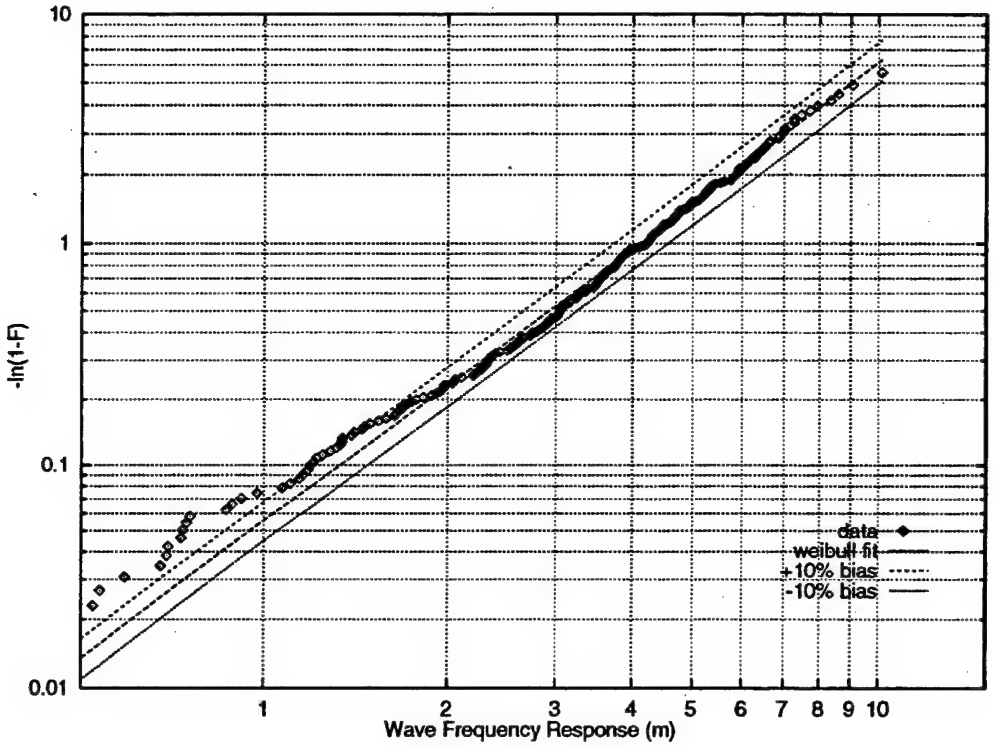

Figure 4: Distribution of local peaks, wave-frequency component.

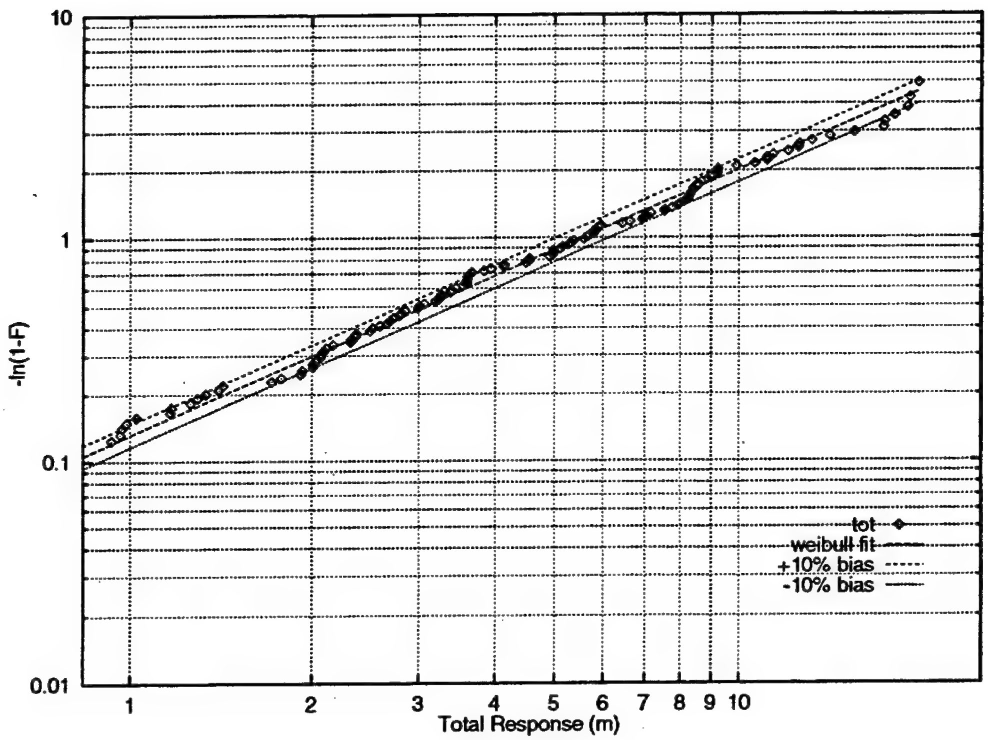

Figure 5: Distribution of local peaks, total response.

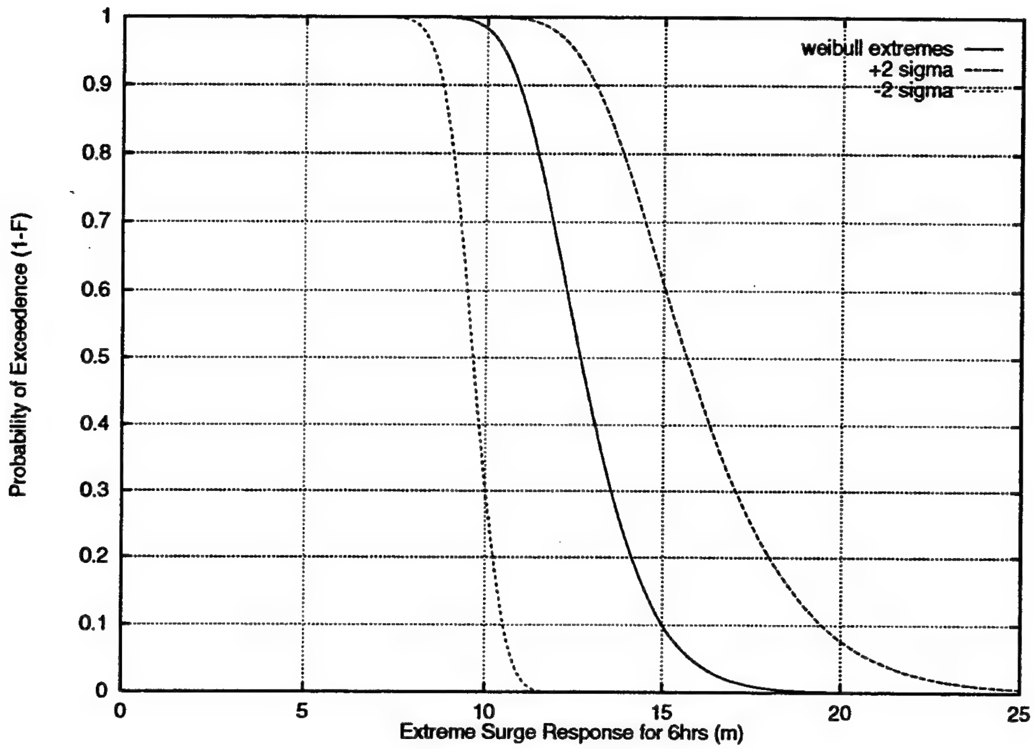

Figure 6: Predicted distribution of 6-hour extreme, surge component.

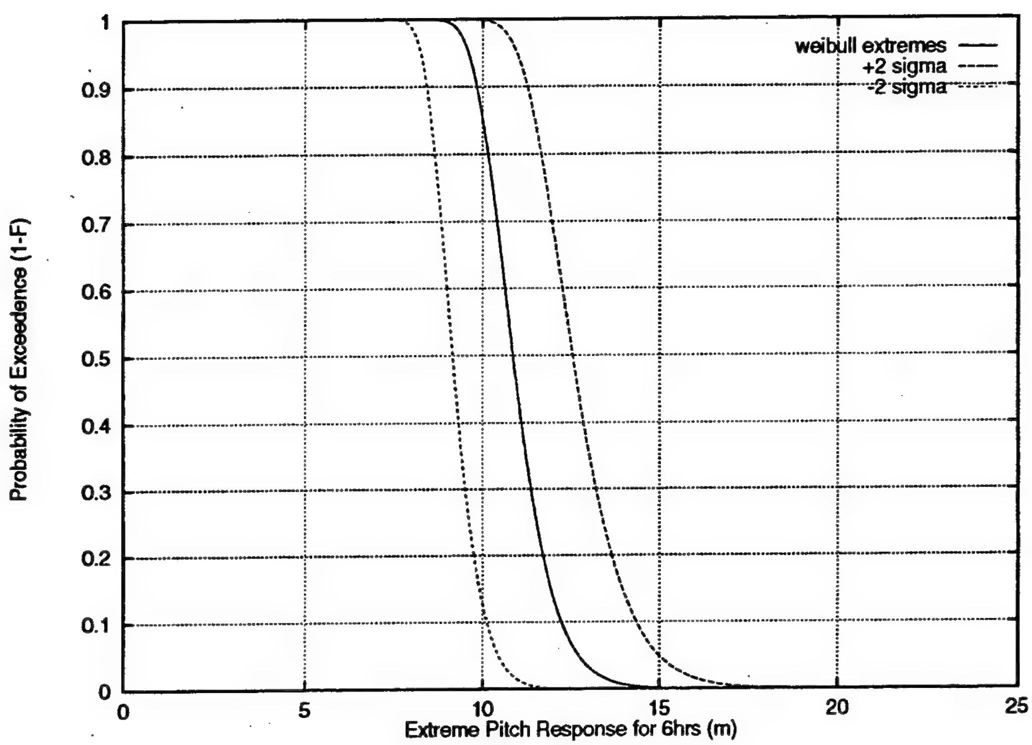

Figure 7: Predicted distribution of 6-hour extreme, pitch component.

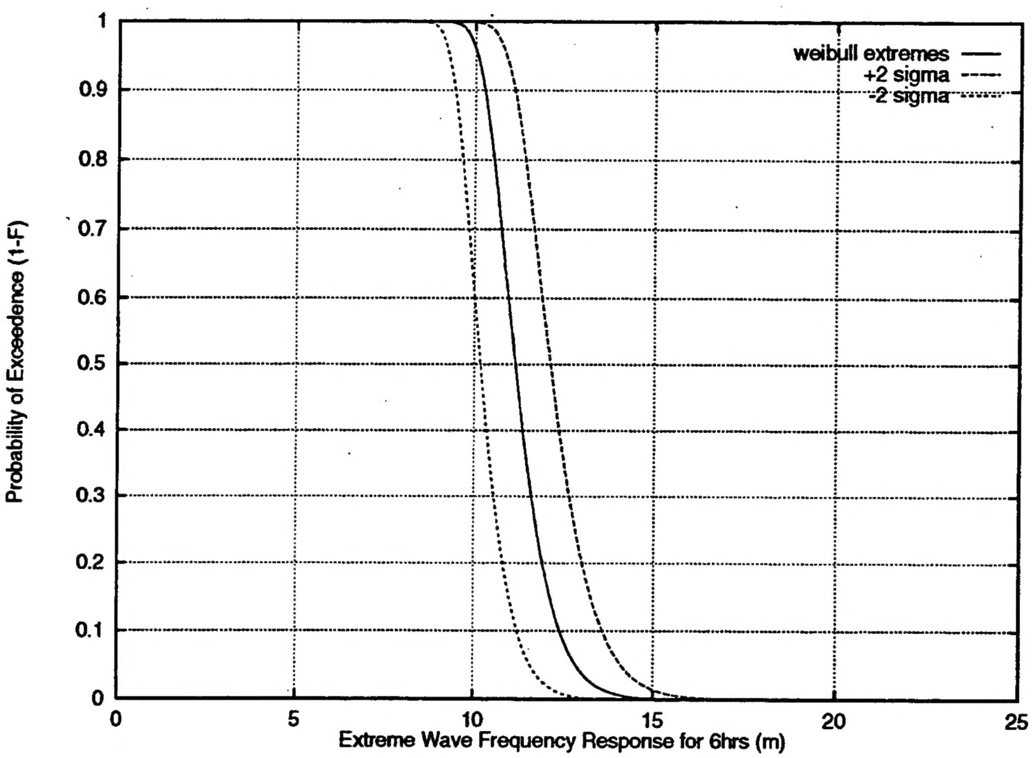

Figure 8: Predicted distribution of 6-hour extreme, wave-frequency component.

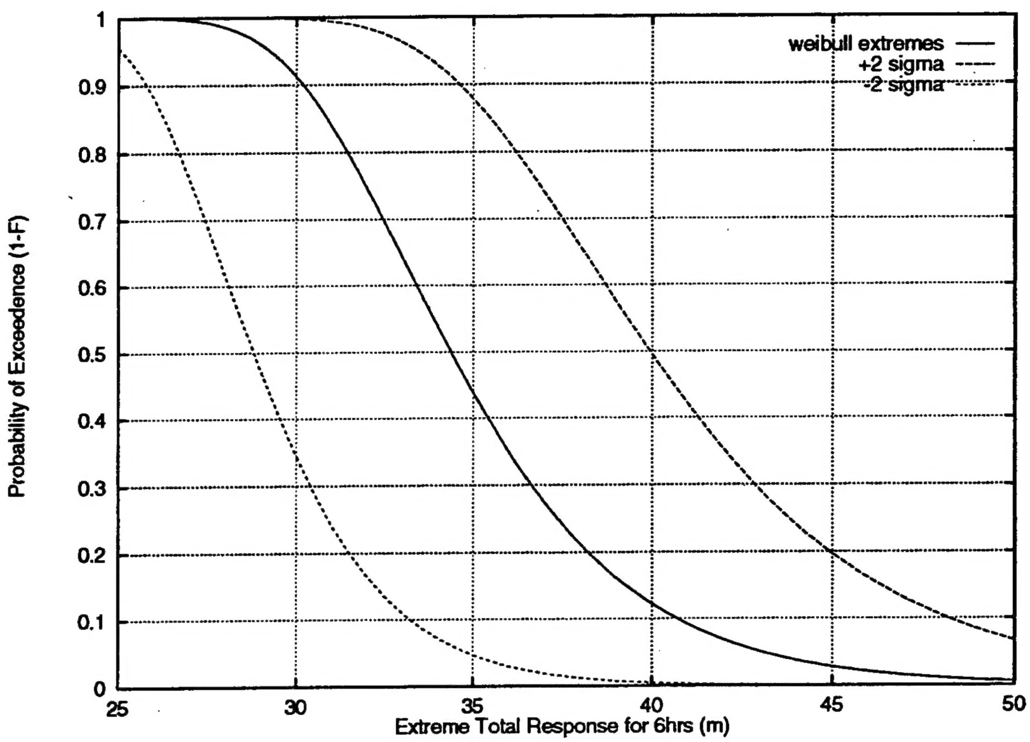

Figure 9: Predicted distribution of 6-hour extreme, total response.

6 References

Efron, B. and R.J. Tibshirani (1993). *An introduction to the bootstrap,* Chapman & Hall.

Engebretsen, K. and S.R. Winterstein (1998). *Probability-based design loads and responses of floating structures: extreme slow drift motion and tether tension,* Rept. RMS–30, Reliability of Marine Structures Program, Dept. of Civ. & Environ. Eng., Stanford University.

Kashef, T. and S.R. Winterstein (1998). *Moment-based probability modelling and extreme response estimation: The FITS routine,* Rept. RMS–31, Reliability of Marine Structures Program, Dept. of Civ. & Environ. Eng., Stanford University.

Madsen, H.O., S. Krenk, and N.C. Lind (1986). *Methods of structural safety,* Prentice-Hall, Inc., New Jersey.

Ude, T.C. and S.R. Winterstein (1996). *IFORM: An inverse-FORM routine to estimate response levels with specified return periods,* Rept. TN–3, Reliability of Marine Structures Program, Civil Eng. Dept., Stanford University.

Ude, T.C., S. Kumar and S.R. Winterstein (1995). *TFPOP 2.0: Stochastic response analysis of floating structures under wind, current, and second-order wave loads,* Rept. RMS–18, Reliability of Marine Structures Program, Dept. of Civ. Eng., Stanford University.

Winterstein, S.R. and K. Engebretsen (1998). Reliability-based prediction of design loads and responses for floating ocean structures. *Proc., 17th Intl. Offshore Mech. Arctic Eng. Symp.,* ASME, to appear.

Winterstein, S.R., C.H. Lange, and S. Kumar (1994). *FITTING: A subroutine to fit four moment probability distributions to data,* Rept. RMS–14, Reliability of Marine Structures Program, Civil Eng. Dept., Stanford University.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
00-06-1998

2. REPORT DATE
June 1998

3. DATES COVERED *(From - To)*

4. TITLE AND SUBTITLE
Probabilistic Models of Dynamic Response and Bootstrap-Based Estimates of Extremes: The Routine MAXFITS

5a. CONTRACT NUMBER

5b. GRANT NUMBER
N00014-96-1-0641

5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHOR(S)
Ron de Jong
Steven R. Winterstein

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
RMS Group
S. R. Winterstein, C. A. Cornell
Blume Center
Stanford University, CA 94305

8. PERFORMING ORGANIZATION REPORT NUMBER
RMS-34

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research
800 N. Quincy St.
Arlington, VA 22217-4620
Attn: Dr Roshdy Barsoum

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSORING/MONITORING AGENCY REPORT NUMBER

12. DISTRIBUTION AVAILABILITY STATEMENT
Approved for Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This report describes and illustrates the use of the Routine MaxFITS. The routine estimates statistics of extremes corresponding to arbitrary dynamic load or response processes. It estimates extremes from limited-duration time-histories which may arise from physical tests or from simulation. A wide variety of statistics can be estimated for an extreme over an arbitrary duration. The routine also asses the statistical uncertainty associated with these extremal statistics. A number of distribution types and the 4-moment Hermite transformation are employed.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPONE NUMBER *(Include area code)*